Analyzing the shift of Estimated Total Intracranial Volume and Normalize Whole Brain Volume considering the mixed effect of diagnostic group and visit order

1. Introduction

In recent years, how to prevent and cure Alzheimer's disease is becoming a very popular topic. Many scientists and hospitals are trying to build some models to forecast the group of people who are vulnerable to Alzheimer's disease. These experiments could focus on various domains, including gender, age, dominant hand, and brain volume. In the next step, Magnetic Resonance Imaging (MRI) will be a significant method on diagnosing if the patient is demented.

In this study, we will explore the factors that could possibly show difference among the demented and nondemented people, which could help us further understand the situation. We will be using the dataset named INF2178_A4_data regarding the Magnetic Resonance Imaging Comparisons of Demented and Nondemented Adults (sourced from Kaggle). To be specific, not every column of variables will be studied, and our research will be focusing on two main research questions:

Research question1: Can we detect a significant interactive effect between diagnostic group and visit order on the estimated total intracranial volume?

Research question2: How do different diagnostic group and visit order affect the normalize whole brain volume?

2. Data Cleaning and Data Wrangling

In this dataset, we can observe **294** rows and **15** columns. Specifically, 4 of the variables from the dataset are categorical data, 2 variables are for identification, while the rest are numerical data. By viewing the data info and descriptive data, we can see there are 15 missing values in SES and 1 missing MMSE.

To perform further study on the research questions, we will not use all the columns, here are the variables of our choice:

- **Subject ID:** identification of the participant
- **Group:** diagnosis group including demented and nondemented
- **Visit:** the visit order
- **eTIV:** estimated total intracranial volume of the participant
- **nWBV:** normalize whole brain volume of the participant

Since we are not using the variables that contain missing value, we will omit the process of removing these missing values and focus on the variables of our choice.

3. Exploratory Data Analysis

	eTIV	nWBV
count	294.00	294.00
mean	1478.85	0.73
std	176.56	0.03
min	1106.00	0.65
25%	1347.25	0.70
50%	1461.50	0.73
75%	1569.00	0.76
max	2004.00	0.84

Table 1: Descriptive data of interested variables

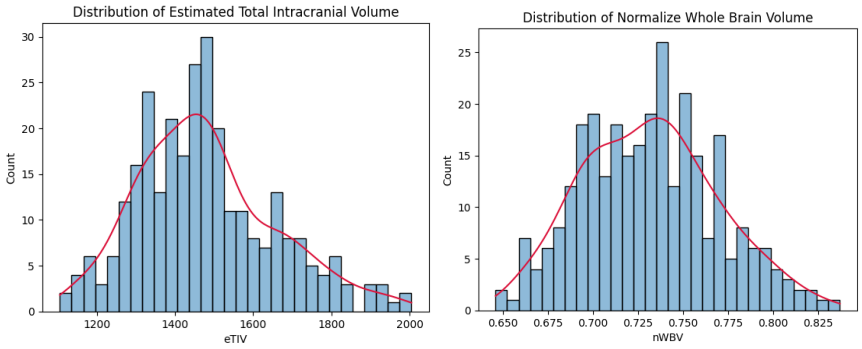

Figure1: Distributions of numerical variables

Given the table 1 and figure 1, The histogram for eTIV shows a distribution that appears slightly right skewed. Most of the data clusters around the middle, as indicated by the mean being near the center of the data's range. For another thing, the histogram for nWBV suggests a more symmetric distribution. Most values cluster around the mean, and there are fewer extreme values.

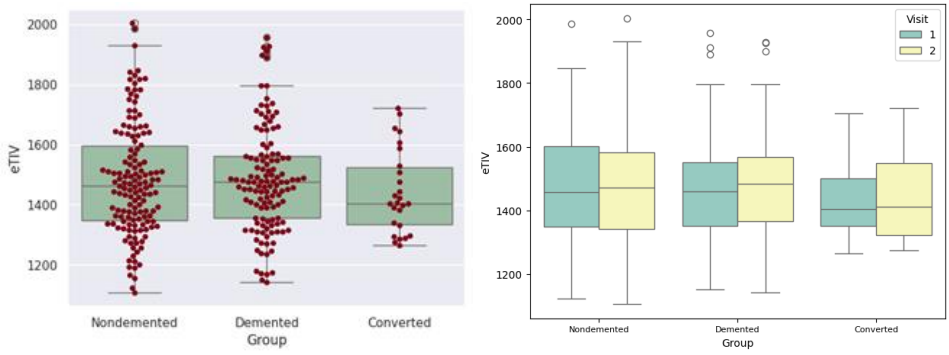

Figure 2: Box plot for eTIV with between subject effect (and within subject effect)

In figure 2, we can see the medians are quite vary from groups, but not vary substantially between visits. The demented groups show most outliers greater than the upper whiskers, while nondemented groups contain little outlier. There is no outlier in converted group.

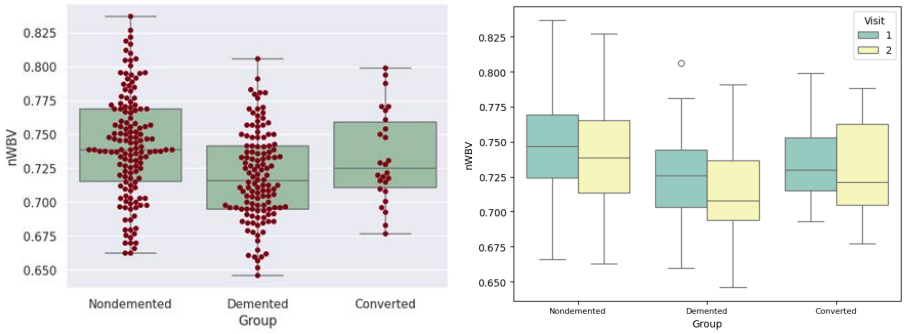

Figure 3: Box plot for nWBV with between subject effect (and within subject effect)

From figure 3 we can see there is only one outlier in demented group. However, the medians are fluctuating among both factors of different groups and visit order.

Figure 4: Power of Test

Given alpha of 0.05, power of 0.91 and effect size 0.7, the power analysis indicates that a sample size of over 44 is necessary to achieve a power of 0.91.

4. Mixed Effect ANOVA

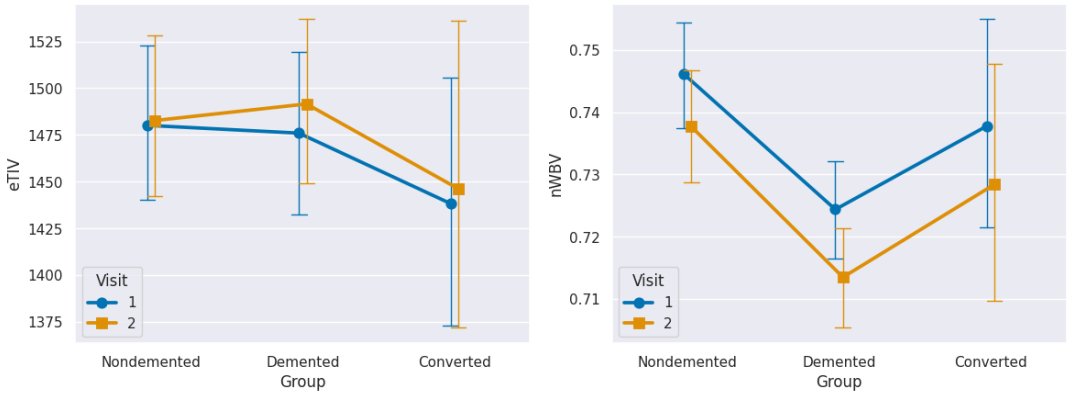

Figure 5: Error bar plot for eTIV and nWBV

As indicated by the result of mixed effect ANOVA, the group factor has an F-value of 0.297 with a p-value of 0.743 for eTIV, 6.712 and 0.02 for nWBV, indicating groups does not have significant

effect in eTIV but does in nWBV. The visit factor has an F-value of 9.225 with a p-value of 0.003 for eTIV and 94.251, <0.001 for nWBV, both variables showing significant differences in visits. Accordingly, the interaction effect has an F-value of 0.831 with a p-value of 0.438 on eTIV and 1.534, 0.219 on nWBV, which refers to no significant interaction between the group and visit.

The trend can also be observed in Figure 5. The error bars are significant, indicating variability within the groups. The nWBV graph demonstrates more pronounced changes between the visits.

5. Assumption Checking

For both numerical variables the mauchly's test of spericity show results of 1, which is a very high p-value means that the test found no violation of the sphericity assumption.

	W	pval	normal
Visit1eTIV	0.97	0.002	False
Visit2eTIV	0.98	0.016	False
Visit1nWBV	0.99	0.37	True
Visit2nWBV	0.99	0.37	True

	W	pval	equal_var
eTIVlevene	0.16	0.69	True
nWBVlevene	0.50	0.48	True

Table 2: Assumption of normality & homogeneity

In table 2, according to the result of Shapiro-Wilk test, the W values are quite high for eTIV, but the p-values are below 0.05, which suggests that the null hypothesis of normal distribution can be rejected. On the other hand, nWBV has p-values greater than 0.05, which means that its normally distributed.

As for the result of homogeneity checking using Levene's test, both variables have relatively low W value, and the p-values are not below 0.05. Therefore, the assumption of homogeneity of variances is met.

6. Conclusion

Now, we can reach to answers for the research questions. Based on the mixed ANOVA analysis for eTIV, we conclude that while there were no significant interaction effects between the diagnostic group and visit order, there was a significant main effect for the visit, indicating that eTIV changes over time. However, this change does not seem to interact with the diagnostic group status of the individuals.

For nWBV, although the data met the assumptions for conducting an ANOVA, the specific effects of the diagnostic group and visit order cannot be determined without the actual ANOVA results. The observed changes in the line graphs suggest there may be an effect, but statistical tests are needed to confirm any conclusions.